

Soc Indic Res (2012) 105:145–160
DOI 10.1007/s11205-010-9770-9

Cross-National Measurement Invariance of the Teacher and Classmate Support Scale

Torbjørn Torsheim · Oddrun Samdal · Mette Rasmussen ·
John Freeman · Robert Griebler · Wolfgang Dür

Accepted: 27 November 2010 / Published online: 16 December 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract The cross-national measurement invariance of the teacher and classmate support scale was assessed in a study of 23202 Grade 8 and 10 students from Austria, Canada, England, Lithuania, Norway, Poland, and Slovenia, participating in the Health Behaviour in School-aged Children (HBSC) 2001/2002 study. A multi-group means and covariance analysis supported configural and metric invariance across countries, but not full scalar equivalence. The composite reliability was adequate and highly consistent across countries. In all seven countries, teacher support showed stronger associations with school satisfaction than did classmate support, with the results being highly consistent across countries. The results indicate that the teacher and classmate support scale may be used in cross-cultural studies that focus on relationships between teacher and classmate support and other constructs. However, the lack of scalar equivalence indicates that direct comparison of the levels support across countries might not be warranted.

Keywords Social support · Measurement · Crosscultural · Self report · HBSC

T. Torsheim (✉)

Department of Psychosocial Science, University of Bergen, Christiesgt 12, 5015 Bergen, Norway
e-mail: torbjoern.torsheim@psych.uib.no

O. Samdal

Department of Health Promotion and Development, University of Bergen, Christiesgt 13,
5015 Bergen, Norway

M. Rasmussen

National Institute of Public Health, University of Southern Denmark, Copenhagen, Denmark

J. Freeman

Faculty of Education, Queen's University, Kingston, ON, Canada

R. Griebler · W. Dür

Ludwig Boltzmann Institute Health Promotion Research, Vienna, Austria

1 Introduction

With increasing international cooperation and policy making in education, there is a growing need for measurements that enable valid analysis of school-related social support across countries. The teacher and classmate support scale (TCMS) (Manger and Olweus 1994; Torsheim et al. 2000) is a brief scale on perceived social support in school, and follows Cobb's (1976) definition of support as information that leads a person to believe she or he is cared for and loved, esteemed and valued or belongs to a network of communication and mutual obligation. The scale consists of a teacher and a classmate subscale, that each includes four Likert-type items about the helpfulness and emotional support from teachers and classmates.

Theoretically, the distinction between teacher and classmate support builds on the notion of informal and formal *support systems* (Cauce et al. 1982). Schools involve both of these systems, with teachers and staff being 'formal' and classmates being 'informal' sources of support. It is important to distinguish between support systems as support from teachers and classmates might serve independent and different functions (Rueger et al. 2010; Wentzel et al. 2010). For example, a student might perceive a high level social support from classmates, but not necessarily a high level of support from teachers. In terms of functional role, teacher support might be related to students coping with academic challenges. Similarly, a central function of perceived classmate support, is provision of social acceptance, and a sense of belongingness. A 'systems approach' to social support in school is compatible with stressor-resource matching theory (Gore and Aseltine 1995) with different systems being relevant for different kinds of demands.

In line with the support systems approach, a confirmatory factor analysis of the TCMS supported a two-factor structure, with moderate correlations between the teacher support factor and the classmate support factor (Torsheim et al. 2000). The study also reported a positive association between teacher and classmate support and school-motivation. In studies that used the TCMS, teacher and classmate support showed concurrent associations with psychological distress (Torsheim and Wold 2001b), somatic complaints (Due et al. 2003; Torsheim and Wold 2001a), and school satisfaction (Danielsen et al. 2009). In longitudinal analysis, social support from teachers and classmates predicted change in distress measured 6 and 12 months later (Torsheim et al. 2003). Thus, the findings on the TCMS converge with a range of other studies in suggesting that school-related social support is of importance to adolescent emotional, social, and academic development (e.g. Dubow et al. 1991; Hamre and Pianta 2005; Kuperminc et al. 2001; Roeser et al. 1996).

Although the TCMS has been used in a number of different countries (Due et al. 2003; Santinello et al. 2009; Torsheim and Wold 2001a), and in comparative studies (Kuntsche et al. 2008; Torsheim et al. 2006), the cross-cultural validity of the scale has received little attention. To draw conclusions about cross-national similarities and differences in social support across comparative studies, some degree of equivalence must exist across the compared groups. For several reasons, however, such equivalence might not be present. For example, at a conceptual level, dimensions of school support might not be equally relevant to all countries—a phenomenon known as concept bias (van de Vijver and Leung 1997). A second threat to cross-cultural validity of the TCMS is lack of measurement invariance. In addressing measurement invariance, it is common to distinguish between configural, metric, and scalar equivalence (e.g. Little 1997; Steenkamp and Baumgartner 1998; Vandenberg 2002). At the most rudimentary level, configural equivalence means that the two-factor structure of school support should be equivalent across cultural groups. If the structure of the measurements differs across groups, comparisons of support or relationships between support and other factor would not be meaningful.

Comparison of social support across culture also requires the metric of the measurements to be equal, so-called *metric equivalence*. That is, a change in teacher support of one unit in one country corresponds to a change of one unit also in other countries. Metric invariance is a prerequisite for directly comparing relationships between social support and other constructs across cultures. When the metric is equivalent, differences and associations observed in country A can be directly compared to differences and associations observed in country B. In contrast, when the metric of social support is non-equivalent, it is difficult to determine whether associations are different due to differences in the underlying metric or true differences due to cultural moderation of associations.

A highly relevant question in cross-cultural research is whether the level of support differs across countries. For example, one might formulate the hypothesis that students in country A have higher levels of teacher support than students in country B, and, subsequently, make an attempt to explain such a difference. This research question requires that the measurement of social support is *scalar equivalent* across the compared countries or cultures. Scalar equivalence means that a score of 3 on teacher support in country A corresponds to a score of 3 in country B. Under assumptions of scalar equivalence, the *level* of social support can be directly compared across countries. However, scalar equivalence is a strong measurement assumption, which might be violated under poor translation or ambiguous wording of items.

To use the TCMS appropriately across countries, there is a need to examine validity and measurement invariance across countries. The objective of the present study was to examine whether or not the TCMS satisfied criteria of configural, metric, and scalar equivalence, and, based on these findings, identify appropriate levels of inference in cross-cultural studies.

2 Method

2.1 Sample

Data were obtained from the Health Behaviour in School-aged Children (HBSC) study of 2001/2002 (Currie et al. 2004). This is a large cross-national study involving 11–16 year-old young people attending school. The study uses a common research protocol that broadly reflects the WHO definition of health as encompassing physical, emotional, and social well being (World Health Organization 1946). Thus the survey includes questions about health status, health complaints, lifestyle health behaviours, family and peer relationships, and school. Findings from each round of the survey have been published in a series of international reports. In each country, the sampling frame was the class with a representative sample of classes selected from a list of all school classes. The desired minimum sample is 1,500 students per age group per country. More detailed information about the sample and the sampling frame can be obtained elsewhere (e.g. Currie et al. 2004). The present sample consisted of 23,202 13 year olds (Grade 8) and 15 year olds (Grade 10) from Austria (2,839 students), Canada (2,720 students), England (3,818 students), Poland (4,235 students), Slovenia (2,455 students), Norway (3,358 students), and Lithuania (3,777 students).

2.2 Procedure

Data collection in each country followed a standardized protocol (Currie et al. 2001). Teachers received instructions on how to administer the survey, and students filled out

questionnaires during ordinary class-hours. Students were informed that participation was voluntary, and that their responses would be anonymous. Students had approximately 45 min to complete the survey. Necessary ethics approval of this protocol was secured on a national basis. In some countries, ethics approval of the protocol was obtained through a regional or national ethics committee. In other countries, such approval was defined as not necessary according to the national regulations and practice at the time of data collection.

2.3 Measurement

The current study was based on a standard English questionnaire. The questionnaire was translated to each country's language(s) by the national research teams. To verify that the integrity of the questionnaire was maintained, independent staff translated the national translation back to English—a procedure popularly known as “back translation” (Harkness 2008). Staff at the HBSC co-ordinating centre in Edinburgh examined the back translation. For validation purposes the present study also used the school satisfaction subscale from Huebner's multidimensional students' life satisfaction scale (Huebner et al. 1998). The teacher and classmate support scale consists of eight items, shown in Appendix. Four of the items refer to perceptions of support from teachers, and four of the items refer to support from classmates. Typical statements were “Other students accept as I am” and “When I need extra help I can get it.” Statements were rated on a 5-point Likert-type agreement scale, ranging from “Strongly agree” (1) to “Strongly disagree” (5). To allow a high score to reflect high support all item scores were reversed prior to analysis.

2.4 Data Analysis

2.4.1 Distributional Assumptions

The observed Likert-scale ratings were treated as continuous scores with interval level measurement properties. The default estimation method in confirmatory factor analysis of continuous data is maximum likelihood (ML). For data that follow a multivariate normal distribution, the ML method has several valuable properties. However, for data that deviate from multivariate normal distribution, ML estimates can be severely biased, and robust maximum likelihood estimation (MLR) is preferable (Finney and DiStefano 2006). Thus, to select an adequate estimator, multivariate normality for the eight observed indicators was tested.

2.4.2 Baseline Model

The two-factor model obtained in a previous validation study was used as a baseline model for testing measurement equivalence (Torsheim et al. 2000). As a necessary first step we assessed the fit of the two-factor model separately in each country. To test the metric properties of each indicator in relation to the latent constructs of teacher and classmate support we assessed the fit of congeneric, tau-equivalent and parallel measurement models. In tau-equivalent measurement models, measurement error terms are allowed to vary across items, but the factor loadings are invariant across items. In parallel measurement models the factor loadings for each observed indicator and the measurement error are invariant across items. Parallel measurement is the implicit assumption in the of simple sum scores.

2.4.3 Testing Equivalence Using Confirmatory Factor Analysis

We used a multiple group means and covariance model (Little 1997) as a framework for testing cross-cultural measurement invariance of the two-factor model. While a number of hypotheses might be relevant in studies of measurement invariance (for an overview see Vandenberg 2002), the current study only included hypotheses relating to configural, metric, and scalar equivalence as these hypotheses were the most relevant to the question of comparability across countries. The first hypothesis to be tested was that the teacher and classmate support scale show configural equivalence. Configural equivalence would be indicated if the two-factor model fitted the data well in all countries. Conversely, if the two-factor model did not account for the item covariance structure in a given country, goodness of fit would deteriorate in that country. In the two-factor baseline model, pattern coefficients, intercepts, and factor variances were estimated freely in each group, but the salient factor loadings had the same structure across groups. To test the internal consistency of the subscales in each of the subsamples, we used Raykov's approach to estimation of lower bound internal consistency for congeneric measures (Raykov 1997). Single item reliability was expressed through the squared factor loading, representing the amount of 'true' variance in each observed indicator.

The second hypothesis to be tested was metric equivalence. In the full metric equivalence model the salient pattern coefficients for each constructs were constrained to be equal across groups. Under full metric equivalence, constraining salient pattern coefficients to equality will be consistent with the observed covariance structure. In contrast, when factor loadings differ across groups, introduction of equality constraints will be inconsistent with the data, thereby resulting in poorer model fit. The amount of change in model fit when equality constraints are imposed might reflect the extent to which the metric differs across groups.

The most restrictive hypothesis in the present study was that of full scalar equivalence. To be able to compare the absolute level of scores across countries, the instrument should display scalar equivalence. To test scalar equivalence, the latent mean structure of teacher and classmate support was included, with factor loadings and intercepts constrained to be equal across groups. The factors would be scalar equivalent if the constraints were consistent with the data.

Assumptions of full measurement equivalence are unrealistic in studies involving groups with different national and cultural backgrounds. A recurring theme in studies of cross-cultural measurement invariance is the degree to which comparability can be achieved under relaxed assumptions. Byrne and colleagues have argued that some degree of comparability can be obtained even under more relaxed assumptions, as long as at least two indicators are invariant (Byrne et al. 1989). In models with a mean structure for the latent variables, partial intercept invariance is essential, since identification of group means require that at least two intercepts are constant across groups.

In the present study, partial measurement invariance was tested based on a sequential release of equality constraints with high modification indices (MI) (Sorbom 1989). Assumptions that are inconsistent with the data are associated with a high MI. In a situation with excess statistical power, such as the present study, the MI is overly sensitive, and tends to indicate modification of even trivial specification error. To evaluate the effect size of the modification, it is necessary to also consider the expected parameter change (EPC) (Kaplan 1989). In the present study, modification indices reflecting expected parameter change (EPC) of weak effect size were not considered, in line with the suggestion in studies of differential item functioning (Jodoin and Gierl 2001). The use of modification

indices in this context primarily serves a diagnostic purpose of detecting items that might compromise the comparisons across countries, to be distinguished from an ad hoc procedure for modifying the overall measurement model.

While assumptions of metric and scalar equivalence can be tested formally through the chi-square difference test, such a test does not convey information about whether the whole model as such fits the data, or whether the model fits the data better than alternative models. In testing the overall fit of the model, model complexity is a potential source of bias (Cheung and Rensvold 2002). Since the scalar equivalence model involves more parameters and more restrictions on these parameters, there might be a bias towards accepting the metric equivalence model, as a function of model complexity, rather than as a function of model appropriateness. A simulation study (Cheung and Rensvold 2002) of fit statistics quantified this bias using a range of fit statistics. The comparative fit index (CFI) showed little bias, and a change in CFI of more than 0.01 point was indicative of a true difference in model fit. In line with Cheung and Rensvold (2002), we used change in CFI as an overall indicator of relative model fit.

2.4.4 Concurrent Validity

As a crude test of concurrent validity, we tested associations with school satisfaction. If the present indicators were to be valid indicators of school support, one would expect that all factors showed association with individual school satisfaction. We expected unique positive association between school satisfaction and teacher support, and between school satisfaction and classmate support.

The clustered sample design introduces dependency between observations from the same class. If such dependency is neglected standard errors tend to be too liberal, with increased type I error rate. To take into account dependency between observations from the same class we used the Mplus COMPLEX option for estimation of all models. Nested model comparisons were done through a chi-square difference test, using the Yuan-Bentler robust maximum likelihood estimation procedure for nonnormal missing data (Yuan and Bentler 2000).

3 Results

3.1 Distributional Assumptions

Table 1 shows univariate (upper part) and multivariate (lower part) skewness and kurtosis. It can be seen from the univariate part that most items showed some departure from normality, as indicated by non-zero skewness and kurtosis. The item “Other students accept me as I am” showed consistent departure from a normal distribution, both in terms of skewness and kurtosis. Mardia’s test of multivariate kurtosis and skewness indicated significant deviation from multivariate normality in all samples. For multivariate skewness the values differed from 2.8 in the Lithuanian sample to 6.9 in the Canadian sample. For multivariate kurtosis the Mardia statistic (Sample estimate—expected value) ranged between 19.74 in the Lithuanian sample and 38.70 in the Canadian sample. Based on these results it was decided to use robust maximum likelihood estimation with the Yuan-Bentler Scaled chi-square test.

Table 1 Tests of univariate and multivariate normality

Variable/label	England	Austria	Canada	Lithuania	Norway	Poland	Slovenia
Univariate test							
Kurtosis							
Extra help from teacher when needed	0.63	0.01	1.66	0.03	0.35	−0.08	0.19
Teacher treat students fairly	−0.40	0.29	−0.23	−0.07	−0.47	−0.82	−0.49
Teacher interest. in student	−0.24	−0.24	−0.16	−0.31	−0.10	−0.42	−0.66
Most teachers are friendly	0.43	0.69	1.29	0.42	1.10	−0.26	−0.54
Student feel down, someone help	−0.06	0.18	−0.02	−0.29	0.45	−0.25	0.56
Students being together	0.27	1.52	0.95	−0.48	0.99	0.55	0.97
Students kind and helpful	−0.50	0.67	0.37	−0.52	0.94	−0.42	0.45
Students accept me	0.69	1.26	1.21	0.60	1.13	1.12	1.49
Skewness							
Extra help from teacher when needed	−0.86	−0.54	−1.17	−0.63	−0.81	−0.69	−0.73
Teacher treat students fairly	−0.42	−0.67	−0.60	−0.63	−0.48	−0.14	−0.43
Teacher interest in student	−0.19	−0.13	−0.48	−0.03	−0.55	−0.42	−0.09
Most teachers are friendly	−0.85	−0.89	−1.11	−0.73	−1.09	−0.67	−0.42
Student feel down, someone help	−0.69	−0.71	−0.63	−0.49	−0.86	−0.64	−0.92
Students being together	−0.56	−1.01	−0.80	−0.20	−0.82	−0.81	−0.87
Students kind and helpful	−0.32	−0.88	−0.80	−0.32	−0.92	−0.47	−0.77
Students accept me	−0.98	−1.10	−1.09	−0.82	−1.06	−0.94	−1.19
Multivariate test							
Kurtosis							
Expected mean	79.98	79.96	79.99	79.97	80.00	79.97	79.99
P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sample value	100.53	105.83	118.69	99.71	109.51	106.37	101.58
Standard deviation	0.39	0.47	0.46	0.39	0.42	0.38	0.48
Skewness							
Expected mean	0.19	0.26	0.28	0.20	0.22	0.17	0.30
P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sample value	4.05	5.66	7.16	3.02	6.64	4.56	5.62
Standard deviation	0.03	0.03	0.04	0.03	0.03	0.02	0.04

3.2 Measurement Model

As a starting point we tested the two-factor model of social support separately in each sample. To assess the metric properties of the items we specified parallel, tau-equivalent and congeneric measurement models. Table 2 shows that in all countries, the congeneric measurement model was the best-fitting model, with CFI-values ranging from 0.968 in the Austrian sample to 0.997 in the Canadian sample. Introducing assumption of tau-equivalence, model fit deteriorated as indicated by strong increase in chi-square and RMSEA, although relative fit indices such as the CFI remained high. Based on these findings the two-factor congeneric measurement model was used as the reference model for multi-group testing of configural, metric and scalar equivalence.

Table 2 Specification of baseline measurement models by subsample

Model/fit indicator	England	Austria	Canada	Lithuania	Norway	Poland	Slovenia
M1 congeneric^a							
CFI	0.991	0.979	0.998	0.988	0.986	0.992	0.994
TLI	0.986	0.968	0.997	0.983	0.979	0.988	0.991
RMSEA	0.026	0.04	0.013	0.032	0.035	0.03	0.022
P-close	1.00	0.985	1.00	1.00	1.00	1.00	1.00
Robust chi-square	69.116	105.192	27.351	90.315	97.112	91.264	41.536
M2 tau-equivalent^b							
CFI	0.966	0.937	0.968	0.958	0.962	0.974	0.961
TLI	0.962	0.93	0.964	0.953	0.958	0.97	0.957
RMSEA	0.043	0.06	0.045	0.052	0.049	0.047	0.047
P-close	0.974	0.006	0.897	0.300	0.58	0.818	0.714
Robust chi-square	204.924	277.059	160.789	276.243	227.892	258.714	163.169
M3 parallel^c							
CFI	0.948	0.916	0.932	0.935	0.914	0.948	0.913
TLI	0.953	0.924	0.938	0.941	0.923	0.953	0.921
RMSEA	0.048	0.062	0.059	0.058	0.067	0.059	0.064
P-close	0.687	0.000	0.006	0.004	0.000	0.000	0.000
Robust chi-square	308.788	367.163	322.2	422.903	493.748	493.208	342.674

^a Degrees of freedom = 19^b Degrees of freedom = 25^c Degrees of freedom = 31

3.3 Measurement Invariance

Table 3 shows that in a multi-group test of configural equivalence, the two-factor solution (model 1) obtained a high fit as indicated by the root mean squared error of approximation (RMSEA) of 0.02 and comparative fit index (CFI) of 0.99. When constraining factor loadings to equality across groups (model 2), the model fit remained strong, consistent with the assumption of metric equivalence across countries. When the intercepts of each observed indicator were constrained to equality (model 3), model fit deteriorated to a CFI of 0.92, suggesting that the assumption of full scalar equivalence was inconsistent with the data. Partial scalar equivalence (model 4) was tested by sequentially releasing cross-group parameter constraints. The intercept of the item “My teachers are interested in me as a person” was associated with a large modification index in the Polish sample, suggesting that at the same level of teacher support, students in Poland tended to have a 0.19 higher score on the item compared to the rest of the groups. The other intercept constraints to be released were “Our teachers treats us fairly” in the Lithuanian and Austrian samples, “Most of the students in my classes are kind and helpful” in the English sample, and “My teachers are interested in me as a person” in the Norwegian sample. The resulting “Partial scalar equivalence model,” achieved an adequate CFI of 0.95, and a RMSEA of 0.051.

Table 4 shows the latent means, variances and covariances for the partial scalar equivalence model. There were in general small differences in the mean level of teacher support and classmate support across countries. For teacher support the level ranged between 3.53 in Poland and 3.89 in Canada. For classmate support the differences ranged

Table 3 Model summary for multi-group test of measurement invariance of correlated two-factor model

Model	Robust Chi-square	df	<i>P</i> <	CFI	TLI	RMSEA	RMSEA 90%CI	Pclose
M1 configural EQ	522.72	133	.001	0.99	0.985	0.030	0.027–0.032	1.00
M2 full metric EQ	807.15	169	.001	0.983	0.980	0.034	0.031–0.036	1.00
M3 full scalar EQ	3,297.62	205	.001	0.917	0.921	0.068	0.066–0.070	0.00
M4 partial scalar EQ	1,912.96	200	.001	0.954	0.955	0.051	0.049–0.053	0.24

between 3.34 in Lithuania and 3.97 in Austria. The covariance between teacher support and classmate support was fairly consistent across countries. The estimated composite reliability was also fairly consistent across countries. In general, the reliability tended to be higher for the teacher subscale, with coefficients in the range of 0.78 (Slovenian sample) to 0.85 (Polish sample). For the classmate support subscale the reliability ranged from 0.70 (English sample) to 0.78 (Polish sample). The single item reliability showed that the reliability tended to be more homogenous for items in the teacher subscale, with squared correlations in the range 0.50–0.60. For the classmate support subscale the squared correlation was more heterogeneous across indicators, with “Most of the students in my class are kind and helpful” showing the highest squared correlation across countries, ranging from 0.58 (English sample) to 0.69 (Austrian sample).

3.4 Association with Other Outcomes

Table 5 shows the results from a multigroup structural equation model, regressing school satisfaction on teacher and classmate support in each of the countries. Concurrent validity would be indicated if classmate support and teacher support showed expected unique associations with school satisfaction across countries. In line with this hypothesis, higher teacher support and classmate support were associated with higher school satisfaction in all countries. In all countries the standardized regression coefficient was markedly stronger for teacher support, with standardized coefficients ranging from 0.34 (Lithuanian sample) to 0.52 (Norwegian and English sample). For classmate support, coefficients ranged from 0.09 (Slovenian sample) to 0.22 (English sample).

4 Discussion

The measurement properties of the TCMS was favourable, as indicated by a clear two-factor structure, sufficient reliability and independent associations with school satisfaction. The observed findings are highly consistent with previous study (Torsheim et al. 2000), but the current results extend the range of validity to several other countries. The fact that a congeneric, but not tau-equivalent or parallel measurement models, fitted the data well, suggests that the current indicators are not exchangeable.

The estimated internal consistency was within acceptable range, although lower than optimal for the classmate support subscale. The observed range of reliability would be appropriate for group-level inferences, but probably not sufficient for diagnostic or classification properties, for example with the objective of identifying children in a school with low social support. Overall this indicate that the present scale is particularly well suited for use in large social surveys or in surveys that seek to reveal substantive relationships between variables. There are numerous examples of such applications of the scale.

Table 4 Estimates (SE) of model parameters for the two-factor model

	England	Austria	Canada	Lithuania	Norway	Poland	Slovenia
Latent means							
Teacher support	3.64 (0.02)	3.70 (0.02)	3.88 (0.02)	3.60 (0.02)	3.83 (0.03)	3.53 (0.03)	3.56 (0.03)
Classmate support	3.45 (0.02)	3.97 (0.03)	3.53 (0.03)	3.34 (0.02)	3.84 (0.02)	3.58 (0.03)	3.80 (0.03)
Variances							
Teacher support	0.44 (0.02)	0.45 (0.03)	0.47 (0.02)	0.40 (0.02)	0.49 (0.03)	0.66 (0.03)	0.49 (0.02)
Classmate support	0.56 (0.02)	0.56 (0.03)	0.64 (0.03)	0.70 (0.03)	0.59 (0.03)	0.71 (0.03)	0.56 (0.03)
Teacher-classmate covariance	0.23 (0.01)	0.20 (0.02)	0.27 (0.02)	0.20 (0.02)	0.24 (0.02)	0.31 (0.02)	0.20 (0.02)
Composite reliability							
Teacher support	0.80 (0.01)	0.80 (0.01)	0.81 (0.01)	0.78 (0.01)	0.79 (0.01)	0.85 (0.01)	0.78 (0.01)
Classmate support	0.70 (0.01)	0.75 (0.01)	0.78 (0.01)	0.76 (0.01)	0.75 (0.01)	0.78 (0.01)	0.73 (0.01)
Item reliability							
Teacher treat students fairly	0.50 (0.02)	0.52 (0.03)	0.53 (0.02)	0.49 (0.02)	0.51 (0.02)	0.58 (0.02)	0.55 (0.03)
Extra help from teacher when needed	0.45 (0.02)	0.50 (0.03)	0.44 (0.02)	0.51 (0.02)	0.48 (0.02)	0.59 (0.02)	0.52 (0.02)
Teacher interest, in student	0.56 (0.02)	0.46 (0.03)	0.61 (0.02)	0.38 (0.02)	0.51 (0.02)	0.59 (0.02)	0.33 (0.03)
Most teachers are friendly	0.46 (0.02)	0.49 (0.03)	0.49 (0.03)	0.54 (0.02)	0.47 (0.02)	0.60 (0.02)	0.48 (0.03)
Student feel down, someone helps	0.31 (0.02)	0.36 (0.02)	0.34 (0.03)	0.32 (0.02)	0.24 (0.02)	0.38 (0.02)	0.31 (0.02)
Students being together	0.30 (0.02)	0.27 (0.02)	0.46 (0.03)	0.38 (0.02)	0.44 (0.02)	0.43 (0.02)	0.45 (0.03)
Students kind and helpful	0.58 (0.02)	0.69 (0.03)	0.69 (0.02)	0.68 (0.02)	0.68 (0.02)	0.65 (0.02)	0.65 (0.03)
Students accept me	0.29 (0.02)	0.40 (0.02)	0.43 (0.03)	0.40 (0.02)	0.42 (0.02)	0.41 (0.02)	0.24 (0.03)

Table 5 Multi-group regression model of school satisfaction on teacher and classmate support with standardised coefficients (SE)

Predictor	Dependent variable: School satisfaction						
	England	Austria	Canada	Lithuania	Norway	Poland	Slovenia
Teacher support	0.52 (0.02)	0.48 (0.02)	0.44 (0.02)	0.34 (0.02)	0.52 (0.02)	0.47 (0.02)	0.45 (0.02)
Classmate support	0.22 (0.02)	0.13 (0.03)	0.22 (0.02)	0.16 (0.02)	0.17 (0.02)	0.19 (0.02)	0.09 (0.03)
R-squared	0.32 (0.02)	0.25 (0.02)	0.27 (0.02)	0.15 (0.01)	0.33 (0.02)	0.27 (0.01)	0.22 (0.02)

In line with the initial validation study (Torsheim et al. 2000), teacher support showed a stronger association with school satisfaction, which might indicate that teacher support has general functions that overlap with other factors of the psychosocial environment. In line with this interpretation, both factors contributed uniquely to predict school satisfaction, but teacher support was the strongest predictor of school satisfaction in all samples.

A key objective of the present study was to assess cross-national measurement invariance of the teacher and classmate support scale. Confirmatory factor analysis revealed strong convergence across countries, but also identified some areas of caution in comparative analysis. By demonstrating configural and metric equivalence, the teacher and classmate support scale satisfied two essential requirements for making comparisons across countries. Under configural equivalence it seems reasonable to argue that the teacher and classmate support scale measures a comparable phenomenon across these countries. That is, irrespective of whether the sample is from Lithuania or Norway, a two-factor model accounted well for the observed item covariances, and the results were much in line with those observed in previous validation efforts (Torsheim et al. 2000).

When comparing causal model associations across countries, it is essential that the observed associations be on the same unit of measurement. Unless it can be ascertained that estimates are on the same metric it is impossible to determine whether differences in associations between social support and other outcomes reflect true differences, or represent an artefact due to different units of measurement. In particular, the finding of metric invariance implies that comparative research focusing on generalizing associations might be warranted, since the underlying metric is comparable, and the reliability is in a similar range across countries.

Ideally, researchers would like to compare the level of support directly across countries, and to make statements such as: “Students in Country A have a higher level of support than students in country B.” However, full scalar equivalence was not supported by the data, as the goodness of fit was greatly reduced under assumptions of intercept invariance across countries. The change in comparative fit was clearly stronger than would be expected due to increased model complexity (Cheung and Rensvold 2002). The post-hoc modifications identified particular problems with two of the teacher items. In a strict sense, the lack of full scalar equivalence means that we should not assume that the scale of measurement has the exact same origin across countries.

One of the distinguishing features of the present study is the large and representative sample. As one of the strengths of the present study, sample bias is not likely to affect the amount of cross-cultural differences. Still, there might be selection effects within each country. School-based surveys, including this one, tend to recruit students who are present

on the day of survey administration. Students who were absent from school on the day of survey administration were not recruited to the study. Comparatively poorer relations with teachers and classmates might be one of the characteristics of non-participants. It is likely that unequal participation by different groups would bias estimates of variance slightly downwards.

In terms of external validity, there are limitations with regards to generalizing the current findings to other countries. Evidence of measurement invariance only applies to comparisons made among these countries and across these two age groups. The measurement quality in other countries and/or for other age groups cannot be extrapolated from the current study, and metric invariance will need to be established separately.

While cross-national studies tend to emphasize differences, the results from the present study highlight cross-national consistency. In spite of language differences, and other economic and cultural differences, youth from the seven countries showed a highly similar structure in their responses to the classroom as measured by the teacher and classmate support scale. Overall, the study indicates that country stratified analyses of this scale are indeed warranted. The finding of only partial, but not full, scalar equivalence indicates some caution in analysis of classroom-based social support pooled across countries, and in making direct comparisons of the level of support across countries.

Acknowledgments Health Behaviour in School-aged Children (HBSC) is an international study carried out in collaboration with WHO/EURO. The International Coordinator of the 2002 survey was Candace Currie and the Data Bank Manager was Oddrun Samdal. We would like to thank the following Principal Investigators: Oddrun Samdal (Norway), Will Boyce (Canada), Antony Morgan (England), Wolfgang Dür (Austria), Helena Jericek (Slovenia), Apolinaras Zaborskis (Lithuania), and Joanna Mazur (Poland). For details, see <http://www.hbsc.org>.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix

See Tables 6 and 7.

Table 6 Items of the teacher and classmate support scale (TCMS)

1. Our teachers treat us fairly
2. When I need extra help, I can get it
3. My teachers are interested in me as a person
4. Most of my teachers are friendly
5. When a student in my class(es) is feeling down, someone else in class tries to help
6. The students in my class(es) enjoy being together
7. Most of the students in my class(es) are kind and helpful
8. Other students accept me as I am

Items are rated on a five-point Likert type scale ranging from 1 (totally agree) to 5 (totally disagree). All items need to be reversed when computing a mean or summed score

Table 7 Estimated sample covariance and means matrix

Item	TCMS 1	TCMS 2	TCMS 3	TCMS 4	TCMS 5	TCMS 6	TCMS 7	TCMS 8
England (N = 3,816)								
TCMS 1	1.017							
TCMS 2	0.471	0.889						
TCMS 3	0.503	0.469	0.968					
TCMS 4	0.482	0.387	0.499	0.939				
TCMS 5	0.195	0.191	0.230	0.207	1.135			
TCMS 6	0.144	0.128	0.143	0.135	0.285	0.814		
TCMS 7	0.281	0.244	0.274	0.250	0.443	0.385	1.002	
TCMS 8	0.132	0.14	0.182	0.116	0.342	0.286	0.400	0.982
Means	3.227	3.695	3.092	3.563	3.557	3.676	3.131	3.707
Austria (N = 2,834)								
TCMS 1	0.957							
TCMS 2	0.49	1.003						
TCMS 3	0.454	0.542	1.02					
TCMS 4	0.518	0.459	0.450	0.958				
TCMS 5	0.179	0.186	0.197	0.219	1.020			
TCMS 6	0.107	0.124	0.112	0.125	0.274	0.654		
TCMS 7	0.213	0.167	0.187	0.227	0.465	0.316	0.865	
TCMS 8	0.182	0.118	0.182	0.141	0.324	0.245	0.461	0.849
Means	3.616	3.606	3.107	3.816	3.824	4.164	3.950	4.096
Canada (N = 2,708)								
TCMS 1	1.148							
TCMS 2	0.493	0.856						
TCMS 3	0.618	0.497	1.050					
TCMS 4	0.521	0.396	0.538	0.896				
TCMS 5	0.255	0.188	0.265	0.217	1.038			
TCMS 6	0.226	0.181	0.244	0.218	0.344	0.733		
TCMS 7	0.289	0.207	0.293	0.250	0.454	0.457	0.889	
TCMS 8	0.208	0.194	0.233	0.183	0.383	0.344	0.497	0.908
Means	3.450	3.939	3.399	3.782	3.514	3.714	3.56	3.791
Lithuania (N = 3,767)								
TCMS 1	0.904							
TCMS 2	0.46	0.949						
TCMS 3	0.365	0.439	0.928					
TCMS 4	0.472	0.454	0.396	0.855				
TCMS 5	0.172	0.195	0.198	0.225	1.100			
TCMS 6	0.133	0.159	0.183	0.149	0.353	0.998		
TCMS 7	0.179	0.206	0.199	0.218	0.493	0.545	1.092	
TCMS 8	0.125	0.147	0.139	0.148	0.371	0.333	0.504	0.846
Means	3.544	3.582	3.018	3.616	3.494	3.403	3.357	3.741
Norway (N = 3,357)								
TCMS 1	1.247							
TCMS 2	0.578	1.037						

Table 7 continued

Item	TCMS 1	TCMS 2	TCMS 3	TCMS 4	TCMS 5	TCMS 6	TCMS 7	TCMS 8
TCMS 3	0.575	0.542	1.112					
TCMS 4	0.562	0.438	0.509	0.978				
TCMS 5	0.176	0.176	0.259	0.225	1.027			
TCMS 6	0.166	0.184	0.188	0.161	0.282	0.735		
TCMS 7	0.221	0.214	0.246	0.239	0.359	0.438	0.836	
TCMS 8	0.169	0.219	0.212	0.185	0.335	0.338	0.477	0.923
Means	3.386	3.747	3.562	3.954	3.862	3.914	3.874	3.934
Poland (N = 4,225)								
TCMS 1	1.287							
TCMS 2	0.717	1.151						
TCMS 3	0.712	0.698	1.176					
TCMS 4	0.745	0.689	0.713	1.217				
TCMS 5	0.32	0.332	0.34	0.352	1.235			
TCMS 6	0.242	0.233	0.215	0.227	0.365	0.860		
TCMS 7	0.33	0.284	0.3	0.345	0.581	0.523	1.093	
TCMS 8	0.2	0.187	0.199	0.219	0.402	0.365	0.468	0.784
Means	3.143	3.541	3.302	3.484	3.57	3.845	3.5	3.904
Slovenia (N = 2,454)								
TCMS 1	1.173							
TCMS 2	0.591	0.985						
TCMS 3	0.477	0.431	1.162					
TCMS 4	0.611	0.534	0.527	1.235				
TCMS 5	0.199	0.197	0.126	0.192	1.009			
TCMS 6	0.16	0.152	0.142	0.151	0.31	0.761		
TCMS 7	0.223	0.212	0.158	0.227	0.416	0.444	0.860	
TCMS 8	0.15	0.157	0.111	0.126	0.298	0.269	0.337	0.909
Means	3.28	3.625	2.965	3.292	3.886	3.876	3.805	3.926

References

- Byrne, B. M., Shavelson, R. J., & Muthen, B. (1989). Testing for the equivalence of factor covariance and mean structures—The issue of partial measurement invariance. *Psychological Bulletin*, 105(3), 456–466.
- Cauce, A. M., Felner, R. D., & Primavera, J. (1982). Social support in high-risk adolescents: Structural components and adaptive impact. *American Journal of Community Psychology*, 10(4), 417–428.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233–255.
- Cobb, S. (1976). Social support as a moderator of life stress. *Psychosomatic Medicine*, 38(5), 300–314.
- Currie, C., Roberts, C., Morgan, A., Smith, R., Settertobulte, W., Samdal, O., et al. (Eds.). (2004). *Young people's health in context. Health behaviour in school-aged children (HBSC) study: International report from the 2001/2002 survey*. Copenhagen: World Health Organization.
- Currie, C., Samdal, O., Boyce, W., & Smith, R. (2001). *Health behaviour in school-aged children: A WHO cross-national study (HBSC), research protocol for the 2001/2002 survey*. Edinburgh: Child and Adolescent Health Research Unit (CAHRU), University of Edinburgh.

- Danielsen, A. G., Samdal, O., Hetland, J., & Wold, B. (2009). School-related social support and students' perceived life satisfaction. *Journal of Educational Research*, 102(4), 303–318.
- Dubow, E. F., Tisak, J., Causey, D., Hryshko, A., & Reid, G. (1991). A 2-year longitudinal study of stressful life events, social support, and social problem-solving skills—Contribution to childrens behavioral and academic adjustment. *Child Development*, 62(3), 583–599.
- Due, P., Lynch, J., Holstein, B., & Modvig, J. (2003). Socioeconomic health inequalities among a nationally representative sample of Danish adolescents: The role of different types of social relations. *Journal of Epidemiology and Community Health*, 57(9), 692–698.
- Finney, S. J., & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 269–314). Greenwich: Information Age Publishing.
- Gore, S., & Aseltine, R. H. (1995). Protective processes in adolescence- matching stressors with social resources. *American Journal of Community Psychology*, 23(3), 301–327.
- Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development*, 76(5), 949–967.
- Harkness, J. A. (2008). Comparative survey research: Goal and challenges. In E. Leeuw, J. J. Hox, & D. Dillman (Eds.), *International handbook of survey methodology*. New York: Lawrence Erlbaum Associates.
- Huebner, E. S., Laughlin, J. E., Ash, C., & Gilman, R. (1998). Further validation of the multidimensional students' life satisfaction scale. *Journal of Psychoeducational Assessment*, 16(2), 118–134.
- Jodoin, M. G., & Gierl, M. J. (2001). Evaluating type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. *Applied Measurement in Education*, 14(4), 329–349.
- Kaplan, D. (1989). Model modification in covariance structure analysis—Application of the expected parameter change statistic. *Multivariate Behavioral Research*, 24(3), 285–305.
- Kuntsche, E., Overpeck, M., & Dallago, L. (2008). Television viewing, computer use, and a hostile perception of classmates among adolescents from 34 countries. *Swiss Journal of Psychology*, 67(2), 97–106.
- Kuperminc, G. P., Leadbeater, B. J., & Blatt, S. J. (2001). School social climate and individual differences in vulnerability to psychopathology among middle school students. *Journal of School Psychology*, 39(2), 141–159.
- Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, 32(1), 53–76.
- Manger, T., & Olweus, D. (1994). A short-form classroom climate questionnaire for elementary and junior high school students. Mimeo. University of Bergen.
- Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, 21(2), 173–184.
- Roeser, R. W., Midgley, C., & Urdan, T. C. (1996). Perceptions of the school psychological environment and early adolescents' psychological and behavioral functioning in school: The mediating role of goals and belonging. *Journal of Educational Psychology*, 88(3), 408–422.
- Rueger, S. Y., Malecki, C. K., & Demaray, M. K. (2010). Relationship between multiple sources of perceived social support and psychological and academic adjustment in early adolescence: Comparisons across gender. *Journal of Youth and Adolescence*, 39(1), 47–61.
- Santinello, M., Vieno, A., & De Vogli, R. (2009). Primary headache in Italian early adolescents: The role of perceived teacher unfairness. *Headache*, 49(3), 366–374.
- Sorbom, D. (1989). Model modification. *Psychometrika*, 54(3), 371–384.
- Steenkamp, J. B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25(1), 78–90.
- Torsheim, T., Aaroe, L. E., & Wold, B. (2003). School-related stress, social support, and distress: Prospective analysis of reciprocal and multi-level relationships. *Scandinavian Journal of Psychology*, 44(2), 153–159.
- Torsheim, T., Ravens-Sieberer, U., Hetland, J., Valimaa, R., Danielson, M., & Overpeck, M. (2006). Cross-national variation of gender differences in adolescent subjective health in Europe and North America. *Social Science and Medicine*, 62(4), 815–827.
- Torsheim, T., & Wold, B. (2001a). School-related stress, school support, and somatic complaints: A general population study. *Journal of Adolescent Research*, 16(3), 293–303.
- Torsheim, T., & Wold, B. (2001b). School-related stress, support, and subjective health complaints among early adolescents: A multilevel approach. *Journal of Adolescence*, 24(6), 701–713.
- Torsheim, T., Wold, B., & Samdal, O. (2000). The teacher and classmate support scale: Factor structure, test-retest reliability and validity in samples of 13 and 15 year-old adolescents. *School Psychology International*, 21(2), 195–212.

- van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage.
- Vandenberg, R. J. (2002). Toward a further understanding of and improvement in measurement invariance methods and procedures. *Organizational Research Methods*, 5(2), 139–158.
- Wentzel, K. R., Battle, A., Russell, S. L., & Looney, L. B. (2010). Social supports from teachers and peers as predictors of academic and social motivation. *Contemporary Educational Psychology*, 35(3), 193–202.
- World Health Organization. (1946). Preamble to the constitution of the World Health Organization as adopted by the international health conference, New York, 19–22 June 1946; signed on 22 July 1946 by the representatives of 61 States (Official Records of the World Health Organization, no. 2, p 100); and entered into force on 7 April 1948.
- Yuan, K. H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological methodology 2000* (Vol. 30, pp. 165–200). Malden: Blackwell Publishing.